

A Generalization of the Chambolle-Pock Algorithm to Banach Spaces with Applications to Inverse Problems

Thorsten Hohage and Carolin Homann

December 2, 2014

Abstract

For a Hilbert space setting Chambolle and Pock introduced an attractive first-order algorithm which solves a convex optimization problem and its Fenchel dual simultaneously. We present a generalization of this algorithm to Banach spaces. Moreover, under certain conditions we prove strong convergence as well as convergence rates. Due to the generalization the method becomes efficiently applicable for a wider class of problems. This fact makes it particularly interesting for solving ill-posed inverse problems on Banach spaces by Tikhonov regularization or the iteratively regularized Newton-type method, respectively.

1 Introduction

Let X and Y be real Banach spaces and $T : X \rightarrow Y$ a linear, continuous operator, with adjoint $T^* : Y^* \rightarrow X^*$. In this paper we will consider the convex optimization problem

$$\bar{x} = \operatorname{argmin}_{x \in X} (g(Tx) + f(x)) \quad (\mathbf{P}) \quad (1)$$

as well as its Fenchel dual problem

$$\bar{p} = \operatorname{argmax}_{p \in Y^*} (-f^*(T^*p) - g^*(-p)), \quad (\mathbf{D}) \quad (2)$$

where $f : X \rightarrow [0, +\infty)$ and $g : Y \rightarrow [0, +\infty)$ belong to the class $\Gamma(X)$ and $\Gamma(Y)$ of proper, convex and lower semicontinuous (l.s.c.) functions. By $f^* \in \Gamma(X^*)$ and $g^* \in \Gamma(Y^*)$ we denote their conjugate functions. A problem of the form (\mathbf{P}) arises in many applications, such as image deblurring (e.g. the ROF model [24]), sparse signal restoration (e.g. the LASSO problem [29]) and inverse problems. We would like to focus on the last aspect. Namely solving a linear ill-posed problem $Tx = y$ by the Tikhonov-type regularization of the general form

$$x_\alpha = \operatorname{argmin}_{x \in X} S(y; Tx) + \alpha R(x), \quad (3)$$

leads for common choices of the data fidelity functional $S(y; \cdot) \in \Gamma(Y)$ and the penalty term $R \in \Gamma(X)$ to a problem of the form (\mathbf{P}) . Also for a nonlinear Fréchet differentiable operator $T : X \rightarrow Y$, where the solution of the operator equation $Tx = y$ can be recovered by the iteratively regularized Newton-type method (IRNM, see e.g. [16])

$$x_{n+1} = \operatorname{argmin}_{x \in X} S(y; T(x_n) + T'[x_n](x - x_n)) + \alpha_n R(x), \quad (4)$$

we obtain a minimization problem of this kind in every iteration step. In particular if S or R is (up to an exponent) given by a norm $\|\cdot\|_Z$ of a Banach space Z it seems to be natural to choose X respectively Y equal to Z . These special problems are of interest in the current research, see e.g. [17, 25, 26] and references therein.

Also inverse problems with Poisson data, which occur for example in photonic imaging, are a topic of certain interest (cf. [16, 33]). Due to the special form of the appropriate choice of S , here proximal algorithms appear to be particularly suitable for minimizing the corresponding regularization functional.

If X and Y are Hilbert spaces one finds a wide class of first-order proximal algorithms in literature for solving (P), e.g. FISTA [6], ADMM [10], proximal splitting algorithms [13]. Chambolle and Pock introduced the following first-order primal-dual algorithm ([11]), which solves the primal problem (P) and its dual (D) simultaneously:

Algorithm 1 (CP). For suitable choices of $\tau_k, \sigma_k > 0$, $\theta_k \in [0, 1]$, $(x_0, p_0) \in X \times Y$, $\hat{x}_0 := x_0$, set:

$$p_{k+1} = (\sigma_k \partial g^* + I)^{-1} (p_k + \sigma_k T \hat{x}_k) \quad (5)$$

$$x_{k+1} = (\tau_k \partial f + I)^{-1} (x_k - \tau_k T^* p_{k+1}) \quad (6)$$

$$\hat{x}_{k+1} = x_{k+1} + \theta_k (x_{k+1} - x_k) \quad (7)$$

Here ∂f denotes the (set-valued) subdifferential of a function f , which will be defined in section 2. There exists generalizations of this algorithm in order to solve monotone inclusion problems ([7, 30]) and to the case of nonlinear operators T ([31]). Recently, Lorenz and Pock ([19]) proposed a quite general forward-backward algorithm for monotone inclusion problems with CP as a special case.

In [11] there are three different parameter choice rules given, for which strong convergence was proven. Two of them base on the assumption that f and/or g^* satisfy a convex property, which enable to prove convergence rates. In order to speed up the convergence of the algorithm, [21] and [15] discuss efficient preconditioning techniques. Thereby the approach studied in [21] can be seen as a generalization from Hilbert spaces X and Y to spaces of the form $\Upsilon^{\frac{1}{2}}X$ and $\Sigma^{-\frac{1}{2}}Y$ for symmetric, positive definite matrices Υ and Σ , where the dual spaces with respect to standard scalar products are given by $\Upsilon^{-\frac{1}{2}}X$ and $\Sigma^{\frac{1}{2}}Y$, respectively. Motivated by this approach, in this paper we further develop a nonlinear generalization of CP to reflexive, smooth and convex Banach spaces X and Y , where we assume X to be 2-convex and Y to be 2-smooth. For all three variations of CP, introduced in [11], we will prove the same convergence results, including linear convergence for the case that f and g^* satisfy a specific convex property on Banach spaces. Moreover the generalization provides clear benefits regarding the efficiency and the feasibility:

First of all the essential factor affecting the performance of the CP-algorithm is the efficiency of calculating the (well-defined, single-valued) resolvents $(\sigma \partial g^* + I)^{-1}$ and $(\tau \partial f + I)^{-1}$ (cf. [32] addressing problem of non exact resolvents in forward-backward algorithms). By the generalization of CP and of the resolvents, inter alia, we obtain closed forms of these operators for a wider class of functions f and g . Furthermore, there exists a more general set of functions that fulfill the generalized convex property on which the accelerated variations of CP are based on. Moreover, in numerical experiments we obtained faster convergence for appropriate choices of X and Y .

The paper is organized as follows: In the next section we give necessary definitions and results of convex analysis and optimization on Banach spaces. In section 3 we present a generalization of CP to Banach spaces, and prove convergence results for special parameter choice rules. The generalized resolvents, which are included in the algorithm, are the topic of section 4. In order to illustrate the numerical performance of the proposed method we apply it in section 5 to some inverse problems. In

particular we consider a problem with sparse solution and a special phase retrieval problem, given as a nonlinear inverse problem with Poisson data.

2 Preliminaries

The following definitions and results from convex optimization and geometry of Banach spaces can be found e.g. in [4, 12].

For a Banach space Z let Z^* denote its topological dual space. In analogy to the inner product on Hilbert spaces, we write $\langle z, z^* \rangle_Z = z^*(z)$ for $z \in Z$ and $z^* \in Z^*$. Moreover, for a function $h \in \Gamma(Z)$ on a Banach space Z let $\partial h : Z \rightrightarrows Z^*$, $z \mapsto \{z^* \in Z^* \mid \forall u \in Z \langle u - z, z^* \rangle_Z \leq h(u) - h(z)\}$ denote the subdifferential of h . Then \bar{x} is the unique minimizer of $h \in \Gamma(X)$ if and only if $0 \in \partial h(\bar{x})$. Moreover, under certain conditions $\bar{x} \in X$ is a solution to the primal problem **(P)** and $-\bar{p} \in Y^*$ is a solution to the dual problem **(D)** if and only if the optimality conditions (see e.g. [36, Section 2.8])

$$-T^* \bar{p} \in \partial f(\bar{x}), \quad T \bar{x} \in \partial g^*(\bar{p}) \quad (8)$$

hold. Another equivalent formulation is that the pair $(\bar{x}, -\bar{p})$ solves the saddle-point problem

$$\min_{x \in X} \max_{p \in Y^*} -\langle Tx, p \rangle_Y + f(x) - g^*(-p) \quad (\mathbf{S}).$$

Rewriting lines (5) and (6) as

$$\sigma_k T \hat{x}_k \in \sigma_k \partial g^* + p_{k+1} - p_k, \quad -\tau_k T^* p_{k+1} \in \tau_k \partial f(x_{k+1}) + x_{k+1} - x_k$$

we can interpret the CP-algorithm as a fixed point iteration with an over-relaxation step in line (7). The objective value can be expressed by the *partial primal-dual gap*:

$$\mathcal{G}_{B_1 \times B_2}(x, p) := \max_{p' \in B_2} (\langle Tx, p' \rangle_Y - g^*(p') + f(x)) - \min_{x' \in B_1} (\langle Tx', p \rangle_Y - g^*(p) + f(x')),$$

on a bounded subset $B_1 \times B_2 \subset X \times Y^*$.

Resolvents On a Hilbert space Z the operator $(\tau \partial h + I)$ is bijective for any $h \in \Gamma(Z)$ and any $\tau > 0$, i.e. the resolvent $(\tau \partial h + I)^{-1} : Z^* = Z \rightarrow Z$ of h is well-defined and single-valued. More generally Rockafellar proved ([22, Proposition 1]) that on any reflexive Banach space Z the function $(\tau \partial h + J_Z)^{-1} : Z^* \rightarrow Z$, where $J_Z = \partial \Phi$ is the subdifferential of $\Phi(x) = \frac{1}{2} \|x\|_Z^2$, is well-defined and single-valued, as well. Furthermore, as $z = (\tau \partial h + J_Z)^{-1}(u)$ is the unique solution of

$$0 \in \partial \tau h(z) + \partial \Phi(z) - u$$

this generalized resolvent can be rewritten as follows:

$$(\tau \partial h + J_Z)^{-1}(u) = \operatorname{argmin}_{z \in Z} \left(\tau h(z) - \langle z, u \rangle_Z + \frac{1}{2} \|z\|_Z^2 \right). \quad (9)$$

Regularity of Banach spaces We make some assumptions on the regularity of the Banach spaces X and Y . A Banach space Z is said to be r -convex with $r > 1$ if there exists a constant $C > 0$, such that the *modulus of convexity* $\delta_Z : [0, 2] \rightarrow [0, 1]$,

$$\delta_Z(\epsilon) := \inf \left\{ 1 - \left\| \frac{1}{2}(x + u) \right\|_Z \mid \|x\|_Z = \|u\|_Z = 1, \|x - u\|_Z \leq \epsilon \right\}$$

satisfies

$$\delta_Z(\epsilon) \geq C \epsilon^r, \quad \epsilon \in [0, 2].$$

We call Z r -smooth, if the *modulus of smoothness* $\rho_Z : [0, \infty] \rightarrow \mathbb{R}$

$$\rho_Z(\tau) := \frac{1}{2} \sup \{ \|x + u\|_Z - \|x - u\|_Z - 2 \mid \|x\|_Z = 1, \|u\|_Z \leq \tau \}$$

fulfills the inequality $\rho_Z(\tau) \leq G_Z \tau^r$ for any $\tau \in [0, \infty]$ and some constant $G_Z > 0$. In the following we will assume both X and the dual space Y^* to be reflexive, smooth and 2-convex Banach spaces. Because of the Lindenstrauss duality formula the second statement is equivalent to the condition that Y is a reflexive, convex and 2-smooth Banach space.

Duality mapping For $q \in (1, \infty)$ let us introduce the *duality mapping*

$$J_{q,Z} : Z \rightrightarrows Z^*, J_{q,Z}(x) := \{x^* \in Z^* \mid \langle x, x^* \rangle_Z = \|x\|_Z \|x^*\|_{Z^*}, \|x^*\|_{Z^*} = \|x\|_Z^{q-1}\}$$

with respect to the weight function $\phi(t) = t^{q-1}$. If Z is smooth, $J_{q,Z}$ is single-valued. If in addition Z is 2-convex and reflexive, $J_{q,Z}$ is bijective with inverse $J_{q^*,Z^*} : Z^* \rightarrow Z^{**} = Z$, where $q^* \in (1, \infty)$, denotes the conjugate exponent of q , i.e. $\frac{1}{q} + \frac{1}{q^*} = 1$. By the theorem of Asplund (see [3]) $J_{q,Z}$ can be also defined as the subdifferential $\partial\Phi_q$ of $\Phi_q(x) := \frac{1}{q} \|x\|_Z^q$. Thus, for the case $q = 2$ the so-called normalized duality mapping $J_{2,Z}$ coincides with the function J_Z , we introduced in the previous section. Note that the duality mapping is in general nonlinear.

Bregman distance Instead of for the functional $(u, v) \mapsto \|u - v\|_X^2$ we will prove our convergence results with respect to the *Bregman distance*

$$\mathcal{B}_Z(u, x) := \frac{1}{2} \|u\|_Z^2 - \frac{1}{2} \|x\|_Z^2 - \langle u - x, J_Z(x) \rangle_Z$$

with gauge function Φ_2 . Note that \mathcal{B}_Z is not a metric, as symmetry is not fulfilled. Nevertheless, a kind of symmetry with respect to the duals holds true:

$$\mathcal{B}_{Z^*}(J_Z(v), J_Z(x)) = \mathcal{B}_Z(x, v). \quad (10)$$

Moreover, the Bregman distance \mathcal{B}_Z satisfies the following identity

$$\begin{aligned} \mathcal{B}_Z(u, x) + \mathcal{B}_Z(v, u) &= \frac{1}{2} \|v\|_Z^2 - \frac{1}{2} \|x\|_Z^2 - \langle u - x, J_Z(x) \rangle_Z - \langle v - u, J_Z(u) \rangle_Z \\ &= \mathcal{B}_Z(v, x) + \langle v - u, J_Z(x) - J_Z(u) \rangle_Z, \quad x, u, v \in Z. \end{aligned} \quad (11)$$

From (9) we observe that

$$(\tau \partial h + J_Z)^{-1}(u) = \operatorname{argmin}_{z \in Z} (\tau h(z) + \mathcal{B}_Z(z, J_Z^*(u)))$$

is a resolvent with respect to the Bregman distance.

The assumption X and Y^* to be reflexive, smooth and 2-convex Banach spaces provide the following helpful inequalities (see e.g. [8]): There exist positive constants C_X and C_{Y^*} , such that:

$$\mathcal{B}_X(x, u) \geq \frac{C_X}{2} \|x - u\|_X^2 \quad \forall x, u \in X, \quad \text{and} \quad \mathcal{B}_{Y^*}(y^*, p) \geq \frac{C_{Y^*}}{2} \|y^* - p\|_{Y^*}^2 \quad \forall y^*, p \in Y^*. \quad (12)$$

Example 1. Considering the proof of the last inequalities (12), we find that the constant $C_X > 0$ comes from a consequence of the Xu-Roach inequalities ([35]):

$$\frac{1}{2} \|x - u\|_X^2 \geq \frac{1}{2} \|x\|_X^2 - \langle J_X x, u \rangle_X + \frac{C_X}{2} \|u\|_X^2, \quad u, x \in X. \quad (13)$$

For example for $X = l^r$ with $r \in (1, 2]$ this estimate holds for $C_X < r - 1$ as it is shown in [34].

Lemma 2. For any $x, u \in X, y^*, p \in Y^*$ and any positive constant α , we have

$$|\langle T(x - u), p - y^* \rangle_Y| \leq \|T\| \left(\frac{\alpha \min \{\mathcal{B}_X(x, u), \mathcal{B}_X(u, x)\}}{C_X} + \frac{\min \{\mathcal{B}_{Y^*}(p, y^*), \mathcal{B}_{Y^*}(y^*, p)\}}{\alpha C_{Y^*}} \right), \quad (14)$$

where $\|T\| = \max \{\|Tx\|_Y \mid x \in X, \|x\|_X = 1\}$ denotes the operator norm.

Proof. Applying Cauchy-Schwarz's inequality as well as the special case of Young's inequality:

$$ab \leq \frac{\alpha a^2}{2} + \frac{b^2}{2\alpha}, \quad a, b \geq 0 \quad (15)$$

with $a := \|x_k - x_{k-1}\|_X$, and $b := \|p_{k+1} - p_k\|_{Y^*} \in \mathbb{R}$ leads to

$$|\langle T(x - u), p - y^* \rangle_Y| \leq \|T\| \|x - u\|_X \|p - y^*\|_{Y^*} \leq \|T\| \left(\alpha \frac{\|x - u\|_X^2}{2} + \frac{\|p - y^*\|_{Y^*}^2}{2\alpha} \right).$$

Now, the inequalities (12) gives the assertion. \square

3 Algorithms and convergence results

Algorithm 2 (CP-BS). For $(\tau_k, \sigma_k)_{k \in \mathbb{N}} \subseteq (0, \infty)$, $\theta \in [0, 1]$, $(x_0, p_0) \in X \times Y^*$, $\hat{x}_0 := x_0$, set:

$$p_{k+1} := (\sigma_k \partial g^* + J_{Y^*})^{-1} (J_{Y^*}(p_k) + \sigma_k T \hat{x}_k) \quad (16)$$

$$x_{k+1} := (\tau_k \partial f + J_X)^{-1} (J_X(x_k) - \tau_k T^* p_{k+1}) \quad (17)$$

$$\hat{x}_{k+1} := x_{k+1} + \theta_k (x_{k+1} - x_k) \quad (18)$$

Let us assume that there is a solution $(\bar{x}, -\bar{p})$ to the saddle-point problem (S). In analogy to [11] we like to bound the distance of one element of the sequence $(x_k, p_k)_{k \in \mathbb{N}}$ to a solution of (S). For the given general Banach space case we measure this misfit by Bregman distances and define for an arbitrary point $(x, y^*) \in X \times Y^*$

$$\Delta_k(x, y^*) := \frac{\mathcal{B}_{Y^*}(y^*, p_k)}{\sigma_k} + \frac{\mathcal{B}_X(x, x_k)}{\tau_k}.$$

Theorem 3. We choose constant parameters $\sigma_k = \sigma, \tau_k = \tau$ and $\theta_k = 1$ for some σ, τ with $\sqrt{\sigma\tau}\|T\| < \min\{C_X, C_{Y^*}\}$, where C_X, C_{Y^*} are given by (12). Then for algorithm 2 the following assertions hold true:

- The sequence $(x_k, p_k)_{k \in \mathbb{N}}$ remains bounded. More precisely there exists a constant

$$C < \left(1 - \frac{\|T\|^2 \sigma \tau}{C_X C_{Y^*}}\right)^{-1},$$

such that for any $N \in \mathbb{N}$

$$\Delta_N(\bar{x}, \bar{p}) \leq C \Delta_0(\bar{x}, \bar{p}). \quad (19)$$

- The restricted primal-dual gap $\mathcal{G}_{B_1 \times B_2}(x^N, p^N)$ at the mean values $x^N := \frac{1}{N} \sum_{k=1}^N x_k \in X$ and $y^N := \frac{1}{N} \sum_{k=1}^N y_k \in Y^*$ is bounded by

$$D(B_1, B_2) := \frac{1}{N} \sup_{(x, y^*) \in B_1 \times B_2} \Delta_0(y^*, x)$$

for any bounded set $B_1 \times B_2 \in X \times Y^*$. Moreover, for every weak cluster point (\bar{x}, \bar{p}) of the sequence $(x^N, p^N)_{N \in \mathbb{N}}$, $(\bar{x}, -\bar{p})$ solves the saddle-point problem (S).

- If we further assume the Banach spaces X and Y to be finite dimensional, then there exists a solution $(\bar{x}, -\bar{p})$ to the saddle-point problem (S) such that the sequence (x_k, p_k) converges strongly to (\bar{x}, \bar{p}) .

Proof. Using (11) the misfit functional $\Delta_k(x, y^*)$ for some $(x, y^*) \in X \times Y^*$ reads

$$\begin{aligned} \Delta_k(x, y^*) &= \frac{\mathcal{B}_Y(J_{Y^*}(p_k), J_{Y^*}(y^*))}{\sigma} + \frac{\mathcal{B}_{X^*}(J_X(x_k), J_X(x))}{\tau} \\ &= \frac{\mathcal{B}_Y(J_{Y^*}(p_{k+1}), J_{Y^*}(y^*))}{\sigma} - \left\langle \frac{J_{Y^*}(p_k) - J_{Y^*}(p_{k+1})}{\sigma}, y^* - p_{k+1} \right\rangle_Y + \frac{\mathcal{B}_Y(J_{Y^*}(p_k), J_{Y^*}(p_{k+1}))}{\sigma} \\ &\quad + \frac{\mathcal{B}_{X^*}(J_{q,X}(x_{k+1}), J_X(x))}{\tau} - \left\langle x - x_{k+1}, \frac{J_X(x_k) - J_X(x_{k+1})}{\tau} \right\rangle_X + \frac{\mathcal{B}_{X^*}(J_X(x_k), J_X(x_{k+1}))}{\tau}. \end{aligned}$$

The iteration formulas (16) and (17) imply that

$$\frac{1}{\sigma} (J_{Y^*}(p_k) - J_{Y^*}(p_{k+1})) + T \hat{x}_k \in \partial g^*(p_{k+1}) \quad \text{and} \quad \frac{1}{\tau} (J_X(x_k) - J_X(x_{k+1})) - T^* p_{k+1} \in f(x_{k+1}).$$

So by the definition of the subdifferential we obtain:

$$g^*(y^*) - g^*(p_{k+1}) \geq \left\langle \frac{J_{Y^*}(p_k) - J_{Y^*}(p_{k+1})}{\sigma}, y^* - p_{k+1} \right\rangle_Y + \langle T \hat{x}_k, y^* - p_{k+1} \rangle_Y \quad (20)$$

$$f(x) - f(x_{k+1}) \geq \left\langle x - x_{k+1}, \frac{J_X(x_k) - J_X(x_{k+1})}{\tau} \right\rangle_X - \langle T(x - x_{k+1}), p_{k+1} \rangle_X, \quad (21)$$

Using (10) this yields

$$\Delta_k(x, y^*) \geq g^*(p_{k+1}) - g^*(y^*) - \langle T \hat{x}_k, p_{k+1} - y^* \rangle_Y + f(x_{k+1}) - f(x) - \langle T(x_{k+1} - x), -p_{k+1} \rangle_X \quad (22)$$

$$+ \frac{\mathcal{B}_{Y^*}(y^*, p_{k+1})}{\sigma} + \frac{\mathcal{B}_{Y^*}(p_{k+1}, p_k)}{\sigma} + \frac{\mathcal{B}_X(x, x_{k+1})}{\tau} + \frac{\mathcal{B}_X(x_{k+1}, x_k)}{\tau} \quad (23)$$

$$+ \langle T x_{k+1}, p_{k+1} - y^* \rangle_Y - \langle T(x_{k+1} - x), p_{k+1} \rangle_Y + \langle T x_{k+1}, y^* \rangle_Y - \langle T x, p_{k+1} \rangle_Y \quad (24)$$

$$= [\langle T x_{k+1}, y^* \rangle_Y - g^*(y^*) + f(x_{k+1})] - [\langle T x, p_{k+1} \rangle_Y - g^*(p_{k+1}) + f(x)] \quad (25)$$

$$+ \Delta_{k+1}(x, y^*) + \frac{\mathcal{B}_{Y^*}(p_{k+1}, p_k)}{\sigma} + \frac{\mathcal{B}_X(x_{k+1}, x_k)}{\tau} + \langle T(x_{k+1} - \hat{x}_k), p_{k+1} - y^* \rangle_Y. \quad (26)$$

In order to estimate the last summand in (26), we insert (18) with $\theta_k = 1$, and apply Lemma 2 with $\alpha := \left(\frac{\sigma}{\tau}\right)^{\frac{1}{2}} > 0$:

$$\begin{aligned} & \langle T((x_{k+1} - x_k) - (x_k - x_{k-1})), p_{k+1} - y^* \rangle_Y \\ &= \langle T(x_{k+1} - x_k), p_{k+1} - y^* \rangle_Y - \langle T(x_k - x_{k-1}), p_k - y^* \rangle_Y - \langle T(x_k - x_{k-1}), p_{k+1} - p_k \rangle_Y \\ &\geq \langle T(x_{k+1} - x_k), p_{k+1} - y^* \rangle_Y - \langle T(x_k - x_{k-1}), p_k - y^* \rangle_Y \\ &\quad - \frac{\|T\| \sigma^{\frac{1}{2}} \tau^{\frac{1}{2}}}{C_X} \frac{\mathcal{B}_X(x_k, x_{k-1})}{\tau} - \frac{\|T\| \sigma^{\frac{1}{2}} \tau^{\frac{1}{2}}}{C_{Y^*}} \frac{\mathcal{B}_{Y^*}(p_{k+1}, p_k)}{\sigma}. \end{aligned}$$

Thus, we conclude that

$$\begin{aligned} \Delta_k(x, y^*) &\geq [\langle T x_{k+1}, y^* \rangle_Y - g^*(y^*) + f(x_{k+1})] - [\langle T x, p_{k+1} \rangle_Y - g^*(p_{k+1}) + f(x)] \\ &\quad + \Delta_{k+1}(x, y^*) + \left(1 - \frac{\|T\| \sigma^{\frac{1}{2}} \tau^{\frac{1}{2}}}{C_{Y^*}}\right) \frac{\mathcal{B}_{Y^*}(p_{k+1}, p_k)}{\sigma} - \frac{\|T\| \sigma^{\frac{1}{2}} \tau^{\frac{1}{2}}}{C_X} \frac{\mathcal{B}_X(x_k, x_{k-1})}{\tau} \\ &\quad + \frac{\mathcal{B}_X(x_{k+1}, x_k)}{\tau} + \langle T(x_{k+1} - x_k), p_{k+1} - y^* \rangle_Y - \langle T(x_k - x_{k-1}), p_k - y^* \rangle_Y. \end{aligned}$$

Summing from $k = 0$ to $N - 1$ leads to

$$\begin{aligned} & \Delta_0(x, y^*) + |\langle T(x_N - x_{N-1}), p_N - y^* \rangle_Y| \\ &\geq \sum_{k=0}^N [\langle T x_{k+1}, y^* \rangle_Y - g^*(y^*) + f(x_{k+1})] - [\langle T x, p_{k+1} \rangle_Y - g^*(p_{k+1}) + f(x)] \\ &\quad + \Delta_N(x, y^*) + \left(1 - \frac{\|T\| \sigma^{\frac{1}{2}} \tau^{\frac{1}{2}}}{C_{Y^*}}\right) \sum_{k=1}^N \frac{\mathcal{B}_{Y^*}(p_k, p_{k-1})}{\sigma} + \frac{\mathcal{B}_X(x_N, x_{N-1})}{\tau} \\ &\quad + \left(1 - \frac{\|T\| \sigma^{\frac{1}{2}} \tau^{\frac{1}{2}}}{C_X}\right) \sum_{k=1}^{N-1} \frac{\mathcal{B}_X(x_k, x_{k-1})}{\tau}. \end{aligned}$$

Now, applying again Lemma 2 with $\alpha = \frac{C_X}{\|T\| \tau}$ yields

$$|\langle T(x_N - x_{N-1}), p_N - y^* \rangle_Y| \leq \frac{\mathcal{B}_X(x_N, x_{N-1})}{\tau} + \frac{\|T\|^2 \sigma \tau \mathcal{B}_{Y^*}(y^*, p_N)}{C_X C_{Y^*} \sigma}, \quad (27)$$

so that we deduce

$$\begin{aligned} \Delta_0(x, y^*) &\geq \sum_{k=0}^N [\langle T x_{k+1}, y^* \rangle_Y - g^*(y^*) + f(x_{k+1})] - [\langle T x, p_{k+1} \rangle_Y - g^*(p_{k+1}) + f(x)] \\ &\quad + \left(1 - \frac{\|T\|^2 \sigma \tau}{C_X C_{Y^*}}\right) \frac{\mathcal{B}_{Y^*}(y^*, p_N)}{\sigma} + \left(1 - \frac{\|T\| \sigma^{\frac{1}{2}} \tau^{\frac{1}{2}}}{C_{Y^*}}\right) \sum_{k=1}^N \frac{\mathcal{B}_{Y^*}(p_k, p_{k-1})}{\sigma} \\ &\quad + \frac{\mathcal{B}_X(x, x_N)}{\tau} + \left(1 - \frac{\|T\| \sigma^{\frac{1}{2}} \tau^{\frac{1}{2}}}{C_X}\right) \sum_{k=1}^{N-1} \frac{\mathcal{B}_X(x_k, x_{k-1})}{\tau}. \end{aligned} \quad (28)$$

Here, because of the choice $\sigma^{\frac{1}{2}} \tau^{\frac{1}{2}} < \frac{\min(C_X, C_{Y^*})}{\|T\|}$, we obtain only positive coefficients. Moreover, for $(x, y^*) = (\bar{x}, \bar{p})$, where $(\bar{x}, -\bar{p})$ solves the saddle point problem (S), we have $-T^* \bar{p} \in \partial f(\bar{x})$ and

$T\bar{x} \in \partial g^*(\bar{p})$, such that every summand in the first line of (28) is non negative as well:

$$\begin{aligned} & [-\langle Tx_{k+1}, -\bar{p} \rangle_Y - g^*(y^*) + f(x_{k+1})] - [\langle T\bar{x}, p_{k+1} \rangle_Y - g^*(p_{k+1}) + f(\bar{x})] \\ & = f(x_{k+1}) - f(\bar{x}) - \langle x_{k+1} - \bar{x}, -T^*\bar{p} \rangle_X + g^*(p_{k+1}) - g^*(\bar{p}) - \langle T\bar{x}, p_{k+1} - \bar{p} \rangle_Y \geq 0. \end{aligned} \quad (29)$$

This proves the first assertion. The second follows directly along the lines of the corresponding proof in [11], p. 124, where we use

$$\begin{aligned} \mathcal{G}_{B_1 \times B_2}(x^N, p^N) &= \sup_{(x, y^*) \in B_1 \times B_2} [\langle Tx_N, y^* \rangle_Y - g^*(y^*) + f(x_N)] - [\langle Tx, p_N \rangle_Y - g^*(p_N) + f(x)] \\ &\leq \frac{1}{N} \sup_{(x, y^*) \in B_1 \times B_2} \Delta_0(x, y^*) \end{aligned}$$

instead of (16). For the last assertion, which needs the assumption that X and Y are finite dimensional, we apply again the same arguments as in [11], p. 124, to (28) from which we obtain

$$\lim_{k \rightarrow \infty} \Delta_k(\bar{x}, \bar{p}) = 0$$

for a solution $(\bar{x}, -\bar{p})$ to the saddle-point problem (S). This completes the proof. \square

Remark 4. This generalization covers also the preconditioned version of CP proposed in [21]: There X and Y are Banach spaces of the form $X = \Upsilon^{\frac{1}{2}}H_X$ with $\|x\|_X = \|\Upsilon^{-\frac{1}{2}}x\|_{H_X}$ and $Y = \Sigma^{-\frac{1}{2}}H_Y$ with $\|y\|_Y = \|\Sigma^{\frac{1}{2}}y\|_{H_Y}$ for Hilbert spaces H_X, H_Y and symmetric, positive definite matrices Υ and Σ . Considering the dual spaces $X^* = \Upsilon^{-\frac{1}{2}}H_X$ and $Y^* = \Sigma^{\frac{1}{2}}H_Y$ with respect to the scalar product on the corresponding Hilbert spaces, the duality mappings read as

$$J_X(x) = \Upsilon^{-1}x, \quad J_{Y^*}(y) = \Sigma^{-1}y.$$

Due to their linearity line (16) and (17) take the form of update rule (4) in [21]:

$$p_{k+1} = (\sigma_k \Sigma \partial g^* + I)^{-1} (p_k + \sigma_k \Sigma T \hat{x}_k), \quad x_{k+1} = (\tau_k \Upsilon \partial f + I)^{-1} (x_k - \tau_k \Upsilon T^* p_{k+1}).$$

In order to generalize also the accelerated forms of the CP-algorithm, which base on the assumption that f is strongly convex, to Banach spaces, we need a similar property of f . More precisely, in [11] the following consequence of f being strongly convex with modulus $\gamma > 0$ is used:

$$f(u) - f(x) \geq 2f\left(\frac{x+u}{2}\right) - 2f(x) + \frac{\gamma}{2}\|x-u\|_X^2 \geq \langle u-x, x^* \rangle_X + \frac{\gamma}{2}\|x-u\|_X^2, \quad u, x \in X, x^* \in \partial f(x).$$

Accordingly, we assume that there exists a constant $\gamma > 0$ such that f satisfies for any $x, u \in X$ and $x^* \in \partial f(x)$ the following inequality

$$f(u) - f(x) \geq \langle u-x, x^* \rangle_X + \gamma \mathcal{B}_X(u, x). \quad (30)$$

With this definition we can formulate the next convergence result. The case that (30) holds for g^* instead of f follows analogously.

Theorem 5. Assume that f satisfies (30) for some $\gamma > 0$ and choose the parameters $(\sigma_k, \tau_k)_{k \in \mathbb{N}}$, $(\theta_k)_{k \in \mathbb{N}}$ in algorithm 2 as follows:

- $\sigma_0 \tau_0 \|T\|^2 \leq \min\{C_X, C_{Y^*}\}$

- $\theta_k := (1 + \gamma \tau_k)^{-\frac{1}{2}}$, $\tau_{k+1} := \theta_k \tau_k$, $\sigma_{k+1} := \theta_k^{-1} \sigma_k$

Then the sequence $(x_k, p_k)_{k \in \mathbb{N}}$ we receive from the algorithm has the following error bound: For any $\epsilon > 0$ there exists an $N_0 \in \mathbb{N}$ such that

$$\mathcal{B}_X(\bar{x}, x_N) \leq \frac{4 + 4\epsilon}{N^2} \left(\frac{\mathcal{B}_X(\bar{x}, x_0)}{\gamma^2 \tau_0^2} + \frac{\mathcal{B}_{Y^*}(\bar{p}, p_0)}{\gamma^2 \sigma_0 \tau_0} \right), \quad (31)$$

for all $N \geq N_0$.

Proof. We go back to the estimate (22)-(26) where we set $(x, y^*) := (\bar{x}, \bar{p})$ for a solution $(\bar{x}, -\bar{p})$ to the saddle-point problem (S). Assumption (30) applied to $x^* = 1/\tau_k (J_X(x_k) - J_X(x_{k+1})) - T^* p_{k+1} \in \partial f(x_{k+1})$ gives:

$$f(\bar{x}) - f(x_{k+1}) \geq \left\langle \bar{x} - x_{k+1}, \frac{J_X(x_k) - J_X(x_{k+1})}{\tau_k} \right\rangle_X - \langle T(\bar{x} - x_{k+1}), p_{k+1} \rangle_X + \gamma \mathcal{B}_X(\bar{x}, x_{k+1}). \quad (32)$$

Thus replacing (21) by (32) and estimating the expansion in line (25) with the help of (29) leads to the inequality:

$$\begin{aligned} \Delta_k(\bar{x}, \bar{p}) &\geq \left(\gamma + \frac{1}{\tau_k} \right) \mathcal{B}_X(\bar{x}, x_{k+1}) + \frac{\mathcal{B}_{Y^*}(\bar{p}, p_{k+1})}{\sigma_k} + \frac{\mathcal{B}_{Y^*}(p_{k+1}, p_k)}{\sigma_k} + \frac{\mathcal{B}_X(x_{k+1}, x_k)}{\tau_k} \\ &\quad + \langle T(x_{k+1} - \hat{x}_k), p_{k+1} - \bar{p} \rangle_Y. \end{aligned}$$

Now we use Lemma 2 with $\alpha = \frac{C_X}{\|T\| \sigma_k \tau_k}$

$$-\theta_{k-1} \langle T(x_k - x_{k-1}), p_{k+1} - p_k \rangle_Y \geq -\frac{\mathcal{B}_{Y^*}(p_{k+1}, p_k)}{\sigma_k} - \frac{\theta_{k-1}^2 \|T\|^2 \tau_{k-1} \sigma_k}{C_X C_{Y^*}} \frac{\mathcal{B}_X(x_k, x_{k-1})}{\tau_{k-1}},$$

and insert $\hat{x}_k = x_k + \theta_k(x_k - x_{k-1})$ from (18) such that we end up with

$$\begin{aligned} \Delta_k(\bar{x}, \bar{p}) &\geq (1 + \gamma \tau_k) \frac{\tau_{k+1}}{\tau_k} \frac{\mathcal{B}_X(\bar{x}, x_{k+1})}{\tau_{k+1}} + \frac{\sigma_{k+1}}{\sigma_k} \frac{\mathcal{B}_{Y^*}(\bar{p}, p_{k+1})}{\sigma_{k+1}} \\ &\quad + \frac{\mathcal{B}_X(x_{k+1}, x_k)}{\tau_k} - \frac{\theta_{k-1}^2 \|T\|^2 \tau_{k-1} \sigma_k}{C_X C_{Y^*}} \frac{\mathcal{B}_X(x_k, x_{k-1})}{\tau_{k-1}} \\ &\quad + \langle T(x_{k+1} - x_k), p_{k+1} - \bar{p} \rangle_Y - \theta_{k-1} \langle T(x_k - x_{k-1}), p_k - \bar{p} \rangle_Y. \end{aligned}$$

The choice of the parameters ensures that

$$(1 + \gamma \tau_k) \frac{\tau_{k+1}}{\tau_k} = \theta_k^{-1} \geq 1, \quad \frac{\sigma_{k+1}}{\sigma_k} = \theta_k^{-1} \geq 1, \quad \text{and} \quad \frac{\tau_k}{\tau_{k+1}} = \theta_k^{-1} \geq 1 \quad \text{for all } k \in \mathbb{N}.$$

Moreover, because of $\min(C_X, C_{Y^*}) \geq \|T\| \tau_0^{\frac{1}{2}} \sigma_0^{\frac{1}{2}} = \|T\| \tau_k^{\frac{1}{2}} \sigma_k^{\frac{1}{2}}$ for $k \in \mathbb{N}$ we have

$$\frac{1}{\tau_k} \frac{\theta_{k-1}^2 \|T\|^2 \tau_{k-1} \sigma_k}{C_X C_{Y^*}} = \frac{1}{\tau_{k-1}} \frac{\|T\|^2 \tau_k \sigma_k}{C_X C_{Y^*}} \leq \frac{1}{\tau_{k-1}}.$$

Therefore

$$\begin{aligned} \frac{\Delta_k(\bar{x}, \bar{p})}{\tau_k} &\geq \frac{\Delta_{k+1}(\bar{x}, \bar{p})}{\tau_{k+1}} + \frac{\mathcal{B}_X(x_{k+1}, x_k)}{\tau_k^2} - \frac{\mathcal{B}_X(x_k, x_{k-1})}{\tau_{k-1}^2} \\ &\quad + \frac{1}{\tau_k} \langle T(x_{k+1} - x_k), p_{k+1} - \bar{p} \rangle_Y - \frac{1}{\tau_{k-1}} \langle T(x_k - x_{k-1}), p_k - \bar{p} \rangle_Y \end{aligned}$$

holds. Now, summing these inequalities from $k = 0$ to $N - 1$ for some $N > 0$ with $x_{-1} := x_0$ and applying (27) with $\tau = \tau_{N-1}$ leads to

$$\begin{aligned} \frac{\Delta_0(\bar{x}, \bar{p})}{\tau_0} &\geq \frac{\Delta_N(\bar{x}, \bar{p})}{\tau_N} + \frac{\mathcal{B}_X(x_N, x_{N-1})}{\tau_{N-1}^2} + \frac{1}{\tau_{N-1}} \langle T(x_N - x_{N-1}), p_N - \bar{p} \rangle_Y \\ &\geq \frac{\Delta_N(\bar{x}, \bar{p})}{\tau_N} - \frac{\|T\|^2}{C_X C_{Y^*}} \mathcal{B}_{Y^*}(\bar{p}, p_N). \end{aligned}$$

By multiplying by τ_N^2 and using the identity $\tau_N \sigma_N = \tau_0 \sigma_0$ we obtain the following error bound:

$$\frac{\tau_N^2}{\tau_0 \sigma_0} \left(1 - \frac{\|T\|^2}{C_X C_{Y^*}} \tau_0 \sigma_0 \right) \mathcal{B}_{Y^*}(\bar{p}, p_N) + \mathcal{B}_X(\bar{x}, x_N) \leq \tau_N^2 \left(\frac{\mathcal{B}_{Y^*}(\bar{p}, p_0)}{\sigma_0 \tau_0} + \frac{\mathcal{B}_X(\bar{x}, x_0)}{\tau_0^2} \right).$$

Substituting γ by $\frac{\gamma}{2}$ in Lemma 1-2 and Corollary 1 in [11] shows that for any $\epsilon > 0$ there exists a $N_0 \in \mathbb{N}$ (depending on ϵ and $\gamma \tau_0$) with $\tau_N^2 \leq 4(1 + \epsilon)(N \gamma)^{-2}$ for all $N \geq N_0$. This completes the proof. \square

Note that compared to the error estimate in [11, Theorem 2] the error bound (31) is 4 times larger for the generalized version CP-BS. That is due to the fact that in the Hilbert space case also the positive term (29) is bounded from below by $\frac{\gamma}{2} \|x_{k+1} - \bar{x}\|_X^2$. In the considered Banach space setting we obtain $\gamma \mathcal{B}_X(x_{k+1}, \bar{x})$ as a lower bound, while $\gamma \mathcal{B}_X(\bar{x}, x_{k+1})$ would be required in order to prove the same result.

Finally, we will show that under the additional assumption that g^* fulfills (30) for some δ we will achieve linear convergence:

Theorem 6. Assume both f and g^* to satisfy property (30) for some constants $\gamma > 0$ and $\delta > 0$, respectively. Then for a constant parameter choice $\sigma_k = \sigma, \tau_k = \tau, \theta_k = \theta, k \in \mathbb{N}$, with

- $\mu \leq \frac{\sqrt{\gamma} \delta \min\{C_X, C_{Y^*}\}}{\|T\|},$
- $\sigma = \frac{\mu}{\delta}, \quad \tau = \frac{\mu}{\gamma},$
- $\theta \in \left[\frac{1}{1+\mu}, 1 \right],$

the sequence $(x_k, p_k)_{k \in \mathbb{N}}$ we receive from algorithm 2 has the error bound:

$$(1 - \omega) \delta \mathcal{B}_{Y^*}(\bar{p}, p_N) + \gamma \mathcal{B}_X(\bar{x}, x_N) \leq \omega^N (\delta \mathcal{B}_{Y^*}(\bar{p}, p_0) + \gamma \mathcal{B}_X(\bar{x}, x_0)), \quad (33)$$

with $\omega = \frac{1+\theta}{2+\mu} < 1$.

Proof. In analogy to the proof of Theorem 5, we obtain from property (30) of f and g^* a sharper estimate for (22)-(26), where we set $(x, y^*) = (\bar{x}, \bar{y})$: We replace (21) by (32) and (20) by

$$g^*(\bar{p}) - g^*(p_{k+1}) \geq \left\langle \frac{J_{Y^*}(p_k) - J_{Y^*}(p_{k+1})}{\sigma}, \bar{p} - p_{k+1} \right\rangle_Y + \langle T \hat{x}_k, \bar{p} - p_{k+1} \rangle_Y + \delta \mathcal{B}_{Y^*}(\bar{p}, p_{k+1}).$$

This together with (29) leads to

$$\begin{aligned} \Delta_k(\bar{x}, \bar{p}) &\geq \left(\delta + \frac{1}{\sigma} \right) \mathcal{B}_{Y^*}(\bar{p}, p_{k+1}) + \left(\gamma + \frac{1}{\tau} \right) \mathcal{B}_X(\bar{x}, x_{k+1}) + \frac{\mathcal{B}_{Y^*}(p_{k+1}, p_k)}{\sigma} + \frac{\mathcal{B}_X(x_{k+1}, x_k)}{\tau} \\ &\quad + \langle T(x_{k+1} - \hat{x}_k), p_{k+1} - \bar{p} \rangle_Y. \end{aligned} \quad (34)$$

Using (18) and (14) with some $\alpha > 0$ we can estimate the last term in the following way:

$$\begin{aligned}
\langle T(x_{k+1} - \hat{x}_k), p_{k+1} - \bar{p} \rangle_Y &= \langle T(x_{k+1} - x_k), p_{k+1} - \bar{p} \rangle_Y - \omega \langle T(x_k - x_{k-1}), p_k - \bar{p} \rangle_Y \\
&\quad - \omega \langle T(x_k - x_{k-1}), p_{k+1} - p_k \rangle_Y - (\theta - \omega) \langle T(x_k - x_{k-1}), p_{k+1} - \bar{p} \rangle_Y \\
&\geq \langle T(x_{k+1} - x_k), p_{k+1} - \bar{p} \rangle_Y - \omega \langle T(x_k - x_{k-1}), p_k - \bar{p} \rangle_Y \\
&\quad - \omega \|T\| \frac{\mathcal{B}_{Y^*}(p_{k+1}, p_k)}{C_{Y^*} \alpha} - \theta \|T\| \alpha \frac{\mathcal{B}_X(x_k, x_{k-1})}{C_X} \\
&\quad - (\theta - \omega) \|T\| \frac{\mathcal{B}_{Y^*}(\bar{p}, p_{k+1})}{C_{Y^*} \alpha},
\end{aligned}$$

for any $\omega \in [(1 + \mu)^{-1}, \theta]$. Now, we set $\alpha = \omega \left(\frac{\gamma}{\delta}\right)^{\frac{1}{2}}$ such that $\frac{\|T\| \mu \omega}{C_{Y^*} \alpha} \leq \delta = \frac{\mu}{\sigma}$ and $\frac{\mu \|T\| \alpha}{C_X} \leq \omega \gamma$ and multiply inequality (34) by μ :

$$\begin{aligned}
\mu \Delta_k(\bar{x}, \bar{p}) &\geq \left(1 + \mu - \frac{1}{\omega}\right) \mu \Delta_{k+1}(\bar{x}, \bar{p}) + \frac{\mu}{\omega} \Delta_{k+1}(\bar{x}, \bar{p}) + \gamma \mathcal{B}_X(x_{k+1}, x_k) - \theta \omega \gamma \mathcal{B}_X(x_k, x_{k-1}) \\
&\quad + \mu \langle T(x_{k+1} - x_k), p_{k+1} - \bar{p} \rangle_Y - \mu \omega \langle T(x_k - x_{k-1}), p_k - \bar{p} \rangle_Y - \frac{(\theta - \omega) \delta}{\omega} \mathcal{B}_{Y^*}(\bar{p}, p_{k+1}).
\end{aligned} \tag{35}$$

As in [11] we choose

$$\omega = \frac{1 + \theta}{2 + \mu} \geq \frac{1 + \theta}{2 + \frac{\sqrt{\gamma \delta} \min\{C_X, C_{Y^*}\}}{\|T\|}}, \tag{36}$$

in order to ensure that

$$\left(1 + \mu - \frac{1}{\omega}\right) \mu \Delta_{k+1}(\bar{x}, \bar{p}) - \frac{(\theta - \omega) \delta}{\omega} \mathcal{B}_{Y^*}(\bar{p}, p_{k+1}) \geq 0.$$

Thus, multiplying (35) with ω^{-k} and summing from $k = 0$ to $N - 1$ for some $N > 0$ where we set $x_{-1} = x_0$) leads to

$$\mu \Delta_0(\bar{x}, \bar{p}) \geq \omega^{-N} \mu \Delta_N(\bar{x}, \bar{p}) + \omega^{-N+1} \gamma \mathcal{B}_X(x_N, x_{N-1}) + \omega^{-N+1} \mu \langle T(x_N - x_{N-1}), p_N - \bar{p} \rangle_Y.$$

Finally, by using Lemma 2 with $\alpha = (\gamma/\delta)^{\frac{1}{2}}$, we obtain from $\|T\| \mu \alpha \leq \gamma \min\{C_X, C_{Y^*}\}$ as well as $\|T\| \mu / \alpha \leq \delta \min\{C_X, C_{Y^*}\}$:

$$\mu \Delta_0(\bar{x}, \bar{p}) \geq \omega^{-N} \mu \Delta_N(\bar{x}, \bar{p}) - \omega^{-N+1} \delta \mathcal{B}_{Y^*}(\bar{p}, p_N),$$

which completes the proof. \square

Remark 7. Because of (12) we proved in Theorem 5 a convergence rate of $\mathcal{O}\left(\frac{1}{N}\right)$, while Theorem 6 even gives a convergence rate of $\mathcal{O}\left(\omega^{\frac{N}{2}}\right)$, i.e. linear convergence.

Remark 8. The parameter choice rules provided by Theorems 3, 5 and 6 depend on the constants C_X and C_{Y^*} given by (12). That is due to the application of Lemma 2 in the corresponding proofs. Now, if we assume that for a specific application this estimate is only required on bounded domains, the constants C_X and C_{Y^*} might be not optimal. In fact, the numerical experiments indicate that we obtain faster convergence if we relax the parameter choice of σ and τ by replacing the product $C_X C_{Y^*}$ by a value $C \in [C_X C_{Y^*}, 1]$ close to 1.

4 Duality mappings and generalized resolvents

In this section we give some examples of duality mappings and discuss the special generalization of the resolvent in our algorithm.

Duality mappings As shown in [14], the reflexive Banach space l^r with $r \in (1, \infty)$ is $\max\{r, 2\}$ -convex and $\min\{r, 2\}$ -smooth. One easily checks that the same holds true for the weighted sequence space l^r_W with positive weight W and norm $\|x\|_{l^r_W} := \left(\sum_j w_j |x_j|^r \right)^{\frac{1}{r}} = \left\| \left(w_j^{\frac{1}{r}} x_j \right)_j \right\|_{l^r}$. With respect to the l^2 -inner product the dual space is given by $Z^* = l^{r^*}_{W^{-1}}$ where $W^{-1} = (w_j^{-1})$ and we have

$$J_{q, l^r_W} : l^r_W \rightarrow l^{r^*}_{W^{-1}}, \quad J_{q, l^r_W}(x) = \frac{W}{\|x\|_{l^r_W}^{r-q}} |x|^{r-1} \operatorname{sign}(x),$$

which has to be understood componentwise.

In order to model for example “blocky” structured solutions $\bar{x} \in X$ let us consider (a discretization of) a Sobolev space $H^{1,r}(\mathbb{T}^d)$ where $\mathbb{T} = (-\pi, \pi)$ with periodic boundary conditions or equivalently the unit circle S^1 . To define these spaces we introduce the Bessel potential operators $\Lambda_s := (I - \Delta)^{-s}$ by

$$\Lambda_s \phi := \sum_{n \in \mathbb{Z}^d} (1 + |n|^2)^{-s/2} \widehat{\phi}(n) \exp(in \cdot), \quad s \in \mathbb{R},$$

a-priori for $\phi \in C^\infty(\mathbb{T}^d)$ where $\widehat{\phi}(n) := (2\pi)^{-d} \int_{\mathbb{T}^d} \exp(-inx) \phi(x) dx$ denote the Fourier coefficients. Note that $\Lambda_0 = I$ and $\Lambda_s \Lambda_t = \Lambda_{s+t}$ for all $s, t \in \mathbb{R}$. For $s \geq 0$ and $r \in (1, \infty)$ the operators Λ_{-s} have continuous extensions to $L^r(\mathbb{T}^d)$, and so the Sobolev spaces

$$H^{s,r}(\mathbb{T}^d) := \Lambda_{-s} L^r(\mathbb{T}^d) \quad \text{with norms} \quad \|\phi\|_{H^{s,r}(\mathbb{T}^d)} := \|\Lambda_s \phi\|_{L^r(\mathbb{T}^d)}$$

are well defined. Actually, this definition also makes sense for $s < 0$, and we have the duality relation

$$\left(H^{r,s}(\mathbb{T}^d) \right)^* = H^{-s,r^*}(\mathbb{T}^d)$$

for $1/r + 1/r^* = 1$ (see e.g. [28, §13.6]). The normalized duality mapping $J_{H^{s,r}(\mathbb{T}^d)} : H^{s,r}(\mathbb{T}^d) \rightarrow H^{-s,r^*}(\mathbb{T}^d)$ is given by

$$J_{H^{s,r}(\mathbb{T}^d)} = \Lambda_{-s} J_{L^r} \Lambda_s.$$

Recall that for $s \in \mathbb{N}$ the space $H^{s,r}(\mathbb{T}^d)$ coincide with the more commonly used Sobolev spaces

$$W^{s,r}(\mathbb{T}^d) := \left\{ \phi \in L^r(\overline{\mathbb{T}^d}) \mid D^\alpha \phi \in L^r(\mathbb{T}^d) \text{ for all } |\alpha| \leq s \right\},$$

with equivalent norms ([28, §13.6]). $H^{s,r}(\mathbb{T}^d)$ is a separable, reflexive, $\max\{r, 2\}$ -convex and $\min\{r, 2\}$ -smooth Banach space (see e.g. [1], [35]).

In the discrete setting we approximate \mathbb{T} by the grid $\mathbb{T}_N := \frac{\pi}{N} \left\{ -\frac{N}{2}, -\frac{N}{2} + 1, \dots, \frac{N}{2} - 1 \right\}$ for some $N \in 2\mathbb{N}$. The dual grid in Fourier space is $\widehat{\mathbb{T}}_N := \frac{N}{\pi} \mathbb{T}_N$, and the discrete Fourier transform $\mathcal{F}_N := (2\pi N)^{-d} (\exp(ix\xi))_{x \in \mathbb{T}_N^d, \xi \in \widehat{\mathbb{T}}_N^d}$ and its inverse \mathcal{F}_N^{-1} can be implemented by FFT. Hence, the Bessel potentials are approximated by the matrices

$$\Lambda_{s,N} := \mathcal{F}_N^{-1} \operatorname{diag} \left[(1 + |\xi|^2)^{-\frac{s}{2}} : \xi \in \widehat{\mathbb{T}}_N^d \right] \mathcal{F}_N.$$

On the finite dimensional space of grid functions $\phi_N : \mathbb{T}_N^d \rightarrow \mathbb{C}$ we introduce the norms

$$\|\phi_N\|_{H^{s,r}(\mathbb{T}_N^d)} := \|\Lambda_{s,N}\phi_N\|_{L^r(\mathbb{T}_N^d)}.$$

If \mathbb{T}_N^d is replaced by $a\mathbb{T}_N^d$ for some $a > 0$, then $\widehat{\mathbb{T}_N^d}$ has to be replaced by $\frac{1}{a}\widehat{\mathbb{T}_N^d}$.

Generalized resolvents Setting $F = J_X$ the resolvent $(\partial f + J_X)^{-1}$ is obviously closely related to the F -resolvents $(A + F)^{-1}F$ of maximal monotone operators A as used in [18] and studied in [5]. Our focus lies on the evaluation of these operators. The following generalization of Moreau's decomposition (see e.g. [23, Theorem 31.5]) allows us to calculate the generalized resolvent $(\sigma_k g^* + J_{Y^*})^{-1}$ in line (16) without knowledge of g^* . This identity can be also derived from [5, Theorem 7.1], but for the convenience of the reader we present a proof for this special case:

Lemma 9. *For any $g \in \Gamma(Y)$, $\sigma > 0$, and $y \in Y$ the minimization problem $\operatorname{argmin}_{z \in Y} \left(\frac{\sigma}{2} \|z - \frac{y}{\sigma}\|_Y^2 + g(z) \right)$ has a unique solution \bar{y} which is equivalently characterized by $J_Y(y - \sigma \bar{y}) \in \partial g(\bar{y})$. Therefore, the operator $(J_{Y^*} \circ \partial g + \sigma I)^{-1} : Y \rightarrow Y$ is well defined and single valued. Moreover, the following identity holds:*

$$(\sigma \partial g^* + J_{Y^*})^{-1}(y) = J_Y \left(y - \sigma (J_{Y^*} \circ \partial g + \sigma I)^{-1}(y) \right), \quad y \in Y \quad (37)$$

Proof. The first assertion follows from [36, Theorem 2.5.1], [12, Theorem 3.4] and the optimality condition $0 \in \partial \left(\frac{\sigma}{2} \|\bar{y} - \frac{y}{\sigma}\|_Y + g(\bar{y}) \right)$. In order to prove the second assertion, we set

$$\bar{y} = (J_{Y^*} \circ \partial g + \sigma I)^{-1}(y)$$

for some $y \in Y$. Moreover, let $\bar{p} \in Y^*$ be a solution to the minimization problem

$$\bar{p} = \operatorname{argmin}_{p \in Y^*} (\sigma g^*(p) - \langle y, p \rangle_Y + \Phi_{2,Y^*}(p)) = \operatorname{argmax}_{p \in Y^*} \left(-g^*(p) + \left\langle \frac{y}{\sigma}, p \right\rangle_Y - \frac{1}{\sigma} \Phi_{2,Y^*}(p) \right),$$

with $\Phi_{2,Y^*}(p) := \frac{1}{2} \|p\|_{Y^*}^2$. Then \bar{p} can be rewritten as $\bar{p} = (\partial \sigma g^* + J_{Y^*})^{-1}(y)$, cf. (9). Because of

$$\left(\frac{1}{\sigma} \Phi_{2,Y^*} \right)^*(y) = \langle y, \sigma J_Y(y) \rangle_Y - \frac{\sigma^2}{2\sigma} \|J_Y(y)\|_Y^2 = \sigma \Phi_{2,Y}(y)$$

$\bar{y} = \operatorname{argmin}_{z \in Y} \left(\sigma \Phi_{2,Y} \left(z - \frac{y}{\sigma} \right) + g(z) \right)$ is the solution $\bar{y} \in Y$ to the corresponding Fenchel dual problem (cf. (1), (2)). Now, (8) implies $-\bar{p} \in \sigma \partial \Phi_{2,Y} \left(\bar{y} - \frac{y}{\sigma} \right) = \sigma J_Y \left(\bar{y} - \frac{y}{\sigma} \right)$. Thus we end up with

$$-(\partial \sigma g^* + J_{Y^*})^{-1}(y) = -\bar{p} = \sigma J_Y \left((J_{Y^*} \circ \partial g + \sigma I)^{-1}(y) - \frac{y}{\sigma} \right). \quad \square$$

Our algorithm appears to be predestined for the case that f and g are given by Banach space norms

$$f(x) := \frac{1}{2} \|x\|_Z^2, \quad g(y) = \frac{1}{2} \|y - y_0\|_W^2$$

for reflexive, smooth and 2-convex Banach spaces Z and W^* and some $y_0 \in W$. The natural choice of the space X and Y in this case is $X = Z$, $Y = W$. Then, due to the theorem of Asplund and

the generalization of Moreau's decomposition (37) the generalized resolvents of f and $g^*(y^*) = \frac{1}{2} (\|y^*\|_{Y^*}^2 + \langle y_0, y^* \rangle_Y)$ reduce to the corresponding duality mappings:

$$(\tau \partial f + J_X)^{-1}(x^*) = (\tau J_X + J_X)^{-1}(x^*) = \frac{1}{\tau + 1} J_{X^*}(x^*) \quad (38)$$

$$(\sigma \partial g^* + J_{Y^*})^{-1}(y) = J_Y \left(y - \sigma (J_Y J_{Y^*}(\cdot - y_0) + \sigma I)^{-1}(y) \right) = J_Y \left(\frac{y - \sigma y_0}{\sigma + 1} \right). \quad (39)$$

Moreover, as we have

$$f(u) - f(x) - \langle u - x, x^* \rangle_X = \mathcal{B}_X(u, x), \quad g^*(p) - g^*(y^*) - \langle y, p - y^* \rangle_Y = \mathcal{B}_Y(y^*, p).$$

for all $u, x \in X$, $x^* \in \partial f(x) = J_X(x)$ and $y^*, p \in Y^*$, $y = \partial J_{Y^*}(p)$, the functions f and g^* satisfy property (30) for all $\gamma, \delta \in (0, 1]$, respectively.

If $X \neq Z$ or $Y \neq W$, however, a system of nonlinear equations has to be solved in order to evaluate the resolvents in lines (16) and (17). In general for all exponents $r \in (1, \infty)$, these resolvents of

$$f(x) := \frac{1}{r} \|x\|_X^r, \quad g(y) := \frac{1}{r} \|y - y_0\|_Y^r, \quad y_0 \in Y \quad (40)$$

become rather simple:

Corollary 10. For $\sigma, \tau > 0$ and f, g given by (40), we have

$$(\tau \partial f + J_X)^{-1}(x^*) = \frac{1}{\tau \alpha^{r-2} + 1} J_{X^*}(x^*), \quad x^* \in X \quad (41)$$

$$(\sigma \partial g^* + J_{Y^*})^{-1}(y) = \frac{1}{\beta^{r-2} + \sigma} J_Y(y - \sigma y_0), \quad y \in Y, \quad (42)$$

where $\alpha \geq 0$ is the maximal solution of $\tau \alpha^{r-1} + \alpha = \|x^*\|_{X^*}$ and $\beta \geq 0$ the maximal solution of $\beta^{r-1} + \sigma \beta = \|y - \sigma y_0\|_Y$.

Proof. Setting $x = (\tau \partial f + J_X)^{-1}(x^*)$, the identity $\partial f(x) = J_{r,X}(x) = \|x\|_X^{r-2} J_X(x)$ implies that $(\tau \|x\|_X^{r-2} + 1) J_X(x) = x^*$, and thus $\alpha = \|x\|_X = \|J_X(x)\|_{X^*} \geq 0$. Inserting α proves the first assertion. For the second one we set $\tilde{y} := (J_{Y^*} \circ \partial g + \sigma I)^{-1}(y)$. Because of $y - \sigma y_0 = J_{Y^*} \circ \partial g(\tilde{y}) + \sigma(\tilde{y} - y_0) = (\|\tilde{y} - y_0\|_Y^{r-2} + \sigma)(\tilde{y} - y_0)$ we have $\beta = \|\tilde{y} - y_0\|_Y$ and $\tilde{y} = \frac{y - \sigma y_0}{\beta^{r-2} + \sigma}$. Now, by Lemma 9 follows:

$$(\sigma \partial g^* + J_{Y^*})^{-1}(y) = J_Y(y - \sigma \tilde{y}) = J_Y \left(\frac{y - \sigma y_0}{\beta^{r-2} + \sigma} \right) \quad \square.$$

Also other standard choices of f for which the resolvent has a closed form, provide a rather simple form for $(\tau \partial f + J_X)^{-1}$ as well.

Example 11. Consider the indicator function $\chi_C \in \Gamma(X)$ of a closed convex set $C \subset X$:

$$\chi_C(x) = \begin{cases} 0 & x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

Then we obtain for any $x^* \in X^*$ and any positive τ

$$\begin{aligned} (\tau \partial \chi_C + J_X)^{-1}(x^*) &= \operatorname{argmin}_{z \in X} \left(\tau \chi_C(z) - \langle z, x^* \rangle_X + \frac{1}{2} \|z\|_X^2 \right) \\ &= \operatorname{argmin}_{z \in C} \left(\frac{1}{2} \|J_{X^*}(x^*)\|_X^2 - \langle z, J_X(J_{X^*}(x^*)) \rangle_X + \frac{1}{2} \|z\|_X^2 \right) = \pi_C(x^*) \end{aligned}$$

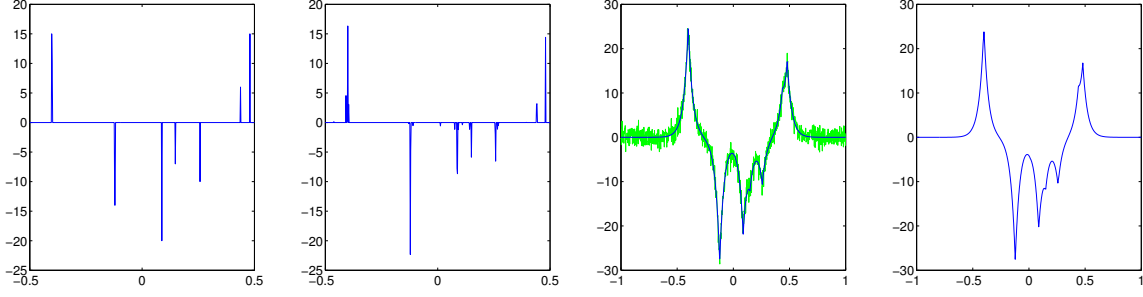


Figure 1: Deconvolution problem with penalty $R(x) = \|x\|_{l^1}$. From left to right: exact solution, reconstruction, exact (blue) and given (green) data, reconstructed data

where $\pi_C : X^* \rightarrow C$ with $\pi_C(x^*) := \underset{z \in C}{\operatorname{argmin}} \mathcal{B}_X(z, J_{X^*}(x^*))$ denotes the generalized projection introduced by Alber [2]. For $f(x) = \|x\|_{l^1}$ the subdifferential is given by

$$\partial f(x) = \begin{cases} \operatorname{sign}(x) & x \neq 0 \\ [-1, 1] & \text{otherwise.} \end{cases}$$

Therefore we have for any $x^* \in X^*$ and any $\tau > 0$

$$(\tau \partial f + J_X)^{-1}(x^*) = J_X(\max\{|x^*| - \tau, 0\} \operatorname{sign}(x^*)).$$

5 Numerical examples

In this section, we will test the performance of the generalized Chambolle-Pock method for linear and nonlinear inverse problems $Tx = y$, i.e. solving (3) or (4). In most examples, X and Y are weighted sequence spaces $l_W^r(I)$ with $r \in (1, \infty)$, countable or finite index sets I , and positive weight W , for which the required operator norm $\|T\|$ is calculated by the power method of Boyd [9]. Also when X is the discrete Sobolev space $H^{s,r}(\mathbb{T}_N^d)$ this method can be applied, since the operators $T : X \rightarrow Y = l_W^r(I)$ and $A := T\Lambda_{-s} : l^r(\mathbb{T}_N^d) \rightarrow Y$ have the same norms. For all versions of the algorithm, we relax the parameter choice of σ and τ according to Remark 8.

First, let us consider a linear ill-posed problem $Tx = y$, with convolution operator

$$T(x) : [-1, 1] \rightarrow \mathbb{R}, \quad T(x)(t) := \int_{-\frac{1}{2}}^{\frac{1}{2}} x(s) k(t-s) ds, \quad k(t) := \exp(-5|t|) \quad (43)$$

and sparse solution $x : [-\frac{1}{2}, \frac{1}{2}] \rightarrow \mathbb{R}$ (see Figure 1). This sparsity constraint is modeled by setting $R(x) = \|x\|_{l^1}$ in (3). Moreover, as instead of the exact data y^\dagger , only data y^δ perturbed by 18 % normal distributed noise is given, we choose $S(y^\delta; y) = \frac{1}{2} \|y^\delta - y\|_{l^2}^2$ as data fidelity functional. According to the properties of the problem, $X = l^r(I_X)$, with $r \in (1, 2]$ and $Y = l^2(I_Y)$, seems to be a good choice. Here, $I_X = \{-\frac{1}{2}, -\frac{1}{2} + \frac{1}{N-1}, \dots, \frac{1}{2}\}$ is the discretization of $[-\frac{1}{2}, \frac{1}{2}]$ and $I_Y = \{-1, -1 + \frac{2}{N-1}, \dots, 1\}$ the of $[-1, 1]$. The discretization of T is the discrete convolution. Now, for $r = 2, 1.75, 1.5, 1.25$ and $\alpha = 5$ we apply the version described in theorem (3) of our algorithm to (3). Inspired by the optimality

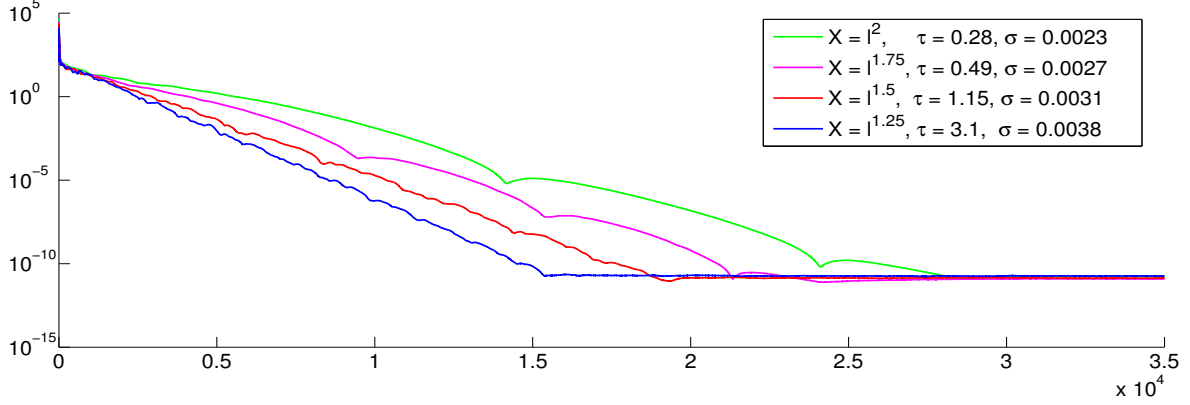


Figure 2: Convergence for the deconvolution problem with penalty $R(x) = \|x\|_{l^1}$. The error $\|x_k - x_\alpha\|_{l^1}$ of the iterates $(x_k)_{k \in \mathbb{N}}$ generated by the algorithm CP-BS1 described in Theorem 3 is plotted over the iteration step k for different choices of X . The parameters τ, σ are chosen optimally.

σ	0.007	0.0023	0.00075
$X = l^2(I_X)$	76476 ($\tau = 0.0915$)	22368 ($\tau = 0.279$)	39418 ($\tau = 0.854$)
$X = l^{1.25}(I_X)$	26271 ($\tau = 1.7$)	16575 ($\tau = 4.8$)	38710 ($\tau = 14.66$)

Table 1: Comparison of CP with $X = l^2(I_X)$ and CP-BS1 with $X = l^{1.25}(I_X)$ for the deconvolution problem with different choices of σ . The table shows the first iterations number k for which $\|x_\alpha - x_k\|_{l^1} \leq 10^{-5}$, averaged over 100 experiments.

condition $T\bar{x} \in \partial g^*(\bar{p})$ with $g^*(p) = S(y^\delta; p)^* = \frac{1}{2}\|p\|_{l^2(I_Y)}^2 + \langle y^\delta, p \rangle_{l^2}$, we pick

$$Tx_0 \in \partial g^*(p_0) = p_0 + y^\delta \Leftrightarrow p_0 = Tx_0 - y^\delta \quad (44)$$

and $x_0 = 0$ as an initial guess. The generalized resolvents are given by (39) and example 11. Figure 2 shows that for experimental optimal chosen parameters σ, τ (according to Remark 8), we obtain faster convergence if r turns 1. As Table 1 illustrates, this holds not only for the optimal parameter choice but also for any other choice of σ . Here, we chose $\tau \in (\sigma^{-1}\|T\|^{-2} - 2^{-6}, \sigma^{-1}\|T\|^{-2})$ for the Hilbert space case $X = l^2(I_X)$, and $\tau \in [\sigma^{-1}\|T\|^{-2}C_1, \sigma^{-1}\|T\|^{-2}C_2]$, with $C_1 = 0.89$, $C_2 = 0.96 \in [0.25, 1]$, for the Banach space case $X = l^{1.25}(I_X)$ (cf. Remark 8). Thus, we conclude that a choice of X , which reflects the properties of the problem best, may provide the fastest convergence.

As a second example, we consider a phase-retrieval problem (see figure 3): a sample of interest is illuminated by a coherent x-ray point source. From intensity measurements $|u^\delta(\cdot, D)|^2$ of the electric field $u : \mathbb{R}^3 \rightarrow \mathbb{C}$, which are taken in the detector plane, orthogonal to the beam at a distance $D > 0$, we want to retrieve information on the refractive index of the sample. More precisely, we are interested in the real phase $\phi : \mathbb{R}^2 \times \{0\} \rightarrow \mathbb{R}$ of the object function $O_\phi(x) = \exp(-ik\phi(x))$ describing the sample, where κ denotes the wavenumber. We assume that $e^{-i\kappa D} u(\cdot, D)$ can be approximated by the so called Fresnel propagator $(P_D O) := \mathcal{F}^{-1} \left(\chi_{-\frac{D}{\kappa}} \cdot (\mathcal{F} O) \right)$ where $\chi_{-c}(t_1, t_2) := \exp(-ic(t_1^2 + t_2^2))$ is a chirp function with parameter $c > 0$. Using the Fresnel scaling theorem we obtain the following forward operator T mapping ϕ to $|u(\cdot, D)|^2$:

$$(T(\phi))(Mt_1, Mt_2) := \frac{1}{M^2} \left| (P_{\frac{D}{M}} O_\phi)(t_1, t_2) \right|^2.$$

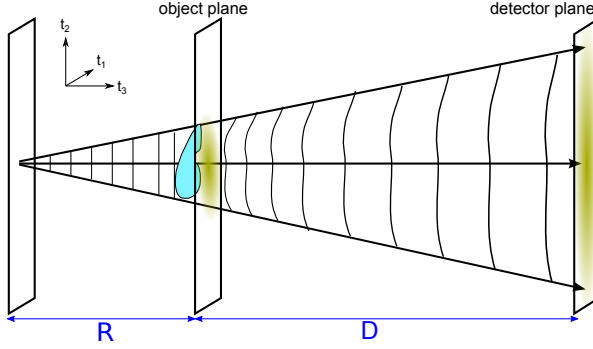


Figure 3: Experimental setup leading to the phase retrieval problem

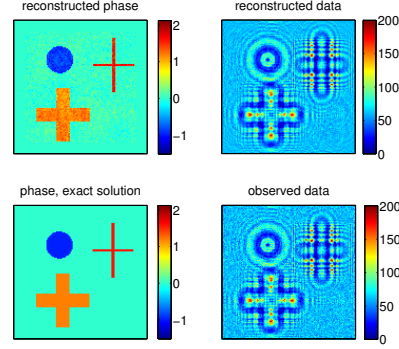


Figure 4: top: reconstructed phase ϕ and corresponding data $T(\phi)$ after 15 IRGN iterations; bottom: exact phase ϕ^\dagger and simulated Poisson distributed diffracton pattern y^δ with $\mathbb{E}[y^\delta] = T(\phi^\dagger)$

Here $M = \frac{R+D}{R}$, where $R > 0$ is the distance between the sample and the source, denotes the geometrical magnification. For a detailed introduction to this problem and phase retrieval problems in general we refer to [20]. T is Fréchet differentiable with

$$T'[\phi](h)(M t_1, M t_2, D) = \frac{2}{M^2} \Re \left(\overline{P_{\frac{D}{M}}(O(\phi))(t_1, t_2)} \left(P_{\frac{D}{M}} O'_{\phi, h} \right)(t_1, t_2) \right)$$

so the IRNM (4) is applicable. This is an example with Poisson data, thus following [16, 33], after the discretization

$$x = (x_j)_{j \in \mathbb{T}_N^2} := \phi(j_1, j_2)_{j \in \mathbb{T}_N^2}, \quad y = (y_j)_{j \in M\mathbb{T}_N^2} := |u(j_1, j_2, D)|^2_{j \in M\mathbb{T}_N^2}$$

we choose (using the convention $0 \ln 0 := 0$):

$$S(y^\delta; y) = \begin{cases} \sum_{j \in M\mathbb{T}_N^2} y_j - y_j^\delta \ln y_j & \text{if } y \geq 0 \text{ and } y_j > 0 \text{ for all } j \text{ with } y_j^\delta > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover, motivated by the weighted least square approximation (cf. [27])

$$S(y^\delta; y) \approx \frac{1}{2} \left\| \left(\frac{y_j^\delta - y_j}{\sqrt{y_j}} \right)_{j \in M\mathbb{T}_N^2} \right\|_{\ell^2}^2,$$

in the $(n+1)$ -th iteration step of the IRNM, we consider the weighted space $Y = \ell_W^2$ with weight $W := (T(\phi_n) + \epsilon)^{-1}$ and $\epsilon = 0.1$. Compared to setting $Y = \ell^2$, this leads to a faster convergence as numerical experiments show. $J_{\ell_W^2}^\rho(y) = Wy$, and $\partial S(y^\delta; y)_j = 1 - \frac{y_j^\delta}{y_j}$ for $(y^\delta, y) \in \text{dom}(S) = \{(y^\delta, y) \mid S(y^\delta; y) < +\infty\}$ imply

$$\left((J_{\ell_W^2}^\rho \circ \partial S(y^\delta, \cdot) + \sigma I)^{-1}(y) \right)_j = \frac{y_j - W_j^{-1}}{2\sigma} + \frac{\sqrt{(y_j - W_j^{-1})^2 + 4\sigma W_j^{-1} y_j^\delta}}{2\sigma}.$$

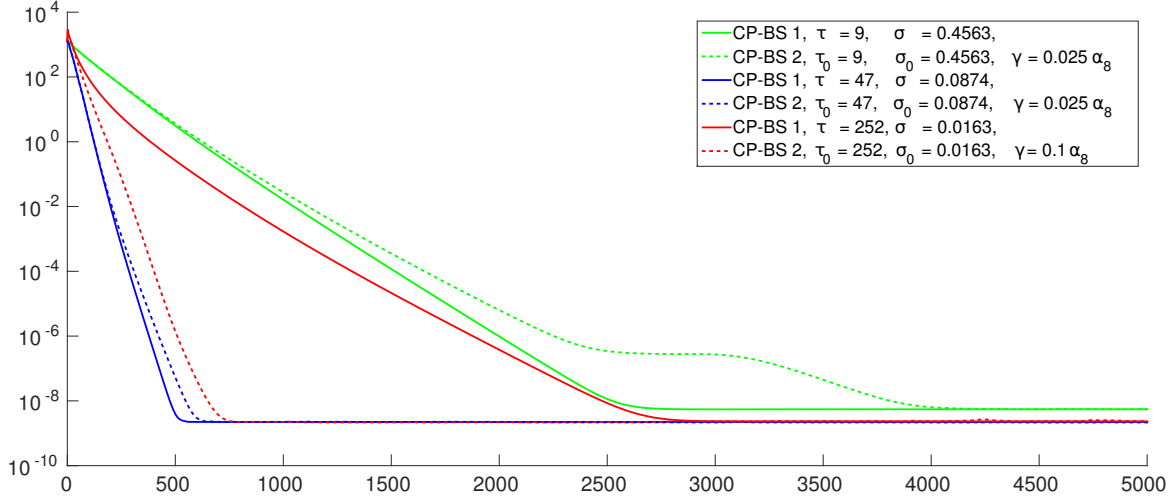


Figure 5: Convergence for the phase retrieval problem with penalty $R(x) = \frac{1}{2}\|x\|_X^2$ for $X = H^{1,1.1}(\mathbb{T}^2)$ at Newton step $n = 8$. The error $\|x_n - x_k\|_X$ of the iterates $(x_k)_{k \in \mathbb{N}}$ of the algorithm CP-BS and the best approximation x_n to the true minimizer of (4) is plotted over the iteration step k . The parameter choice rules are defined by Theorem 3 (CP-BS1, solid) and by Theorem 5 (CP-BS2, dotted), respectively. For given τ (or τ_0), we set $\sigma = 0.96\|T'[\phi_8]\|^{-2}\tau^{-1}$ (σ_0 analogously)

Hence, we obtain the generalized resolvent $(\sigma \partial S(y^\delta, \cdot)^* + I)^{-1}$ by Lemma 9. The "blocky" structured solution is taken into account by setting $X := H^{1,r}(\mathbb{T}^2)$ with $r = 1.1$ and $R(x) = \frac{1}{2}\|x\|_X^2$. Note that although evaluating the generalized resolvent $(\tau \alpha R + J_X)^{-1} = \frac{1}{\tau \alpha + 1} J_{H^{-1,r}} = \frac{1}{\tau \alpha + 1} \Lambda_1 J_{I^*} \Lambda_{-1}$ is more expensive than in the case $X = l^r$, it does not increase the complexity of the algorithm as the evaluations of $T'[\phi]$ and $T'[\phi]^*$ include Fourier transforms as well. Since R satisfies property (30), we can apply the variant CP-BS1 described in Theorem 3 and also the variant CP-BS2 given by Theorem 5. Figure 5 compares both versions in the $n = 8$ -th iteration step of the IRNM, where $\alpha = 0.001$. The solid blue curve belongs to the version CP-BS1 for a optimal parameter choice of τ and σ we found experimentally. Note that for the limit $\gamma \rightarrow 0$ the parameter choice rule of CP-BS2 coincide with the one of CP-BS1. In fact, choosing τ_0 and σ_0 in the same way as τ and σ that corresponds to this blue curve the version CP-BS2 with $\gamma = 0.0025\alpha_8$ gives the same curve. Tuning also the parameters τ_0, σ_0 and γ (reasonable large) in an optimal way, we did not obtain a better convergence result for CP-BS2 than for CP-BS1. However, CP-BS2 converges faster for $\tau = \tau_0$ sufficiently large and adequately chosen γ .

In our last example, we apply the version described by Theorem 6, which we denote as CP-BS3, to the Tikhonov functional

$$\frac{1}{2}\|Tx - y^\delta\|_Y^2 + \frac{\alpha}{2}\|x\|_X^2$$

where T is again the convolution operator (43). We set $\alpha = 1$, $X = l^{1.5}(I_X)$ and $Y = l^2(I_X)$ (see figure 6). Setting $\mu = \frac{C\sqrt{\gamma\delta}}{2\|T\|}$ with $C = 0.98$, we obtain for any choice $\gamma = \delta \in (0, 2]$ the fastest convergence rate. The same rate is also provided by CP-BS1 and CP-BS2 for optimal chosen (initial) parameter. Compared to the first example where more than 15000 iterations were required to satisfy the stopping criterion, here we only need 558 iterations.

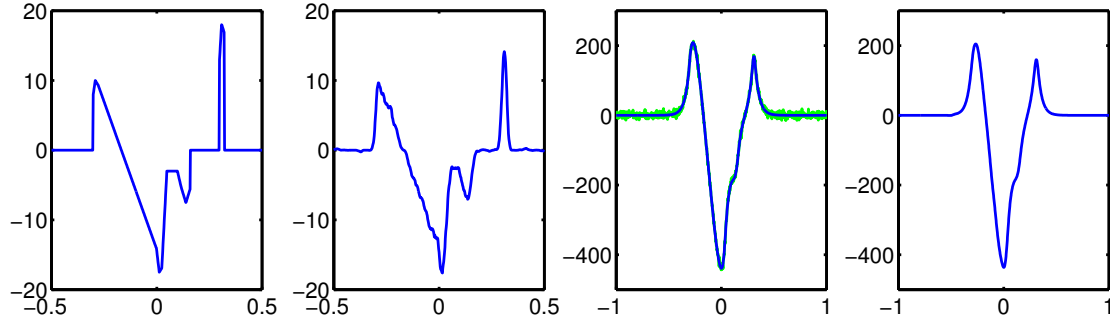


Figure 6: Deconvolution problem with penalty $R(x) = \frac{1}{2}\|x\|_{l^{1.5}}^2$. From left to right: exact solution, reconstruction, exact (blue) and given (green) data, reconstructed data

Acknowledgement

We would like to thank Radu Bot and Russell Luke for helpful discussions. Financial support by DFG through CRC 755, project C2 is gratefully acknowledged.

References

- [1] R. Adams and J. Fournier. *Sobolev Spaces*. Pure and Applied Mathematics. Elsevier Science, 2003.
- [2] Y. I. Alber. Generalized projection operators in Banach spaces: properties and applications. *Functional Differential Equations, Proceedings of the Israel Seminar in Ariel*, 1:1–21, 1994.
- [3] E. Asplund. Positivity of duality mappings. *Bulletin of the American Mathematical Society*, 73(2):200–203, 03 1967.
- [4] V. Barbu and T. Precupanu. *Convexity and Optimization in Banach Spaces*. Mathematics and Its Applications (East European Series). București: D. Reidel Publishing Company, 1986.
- [5] H. H. Bauschke, X. Wang, and L. Yao. General resolvents for monotone operators: characterization and extension, biomedical mathematics: Promising directions. In *in Imaging, Therapy Planning and Inverse Problems*, Medical Physics Publishing, pages 57–74, 2010.
- [6] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Img. Sci.*, 2(1):183–202, Mar. 2009.
- [7] R. I. Boț, E. R. Csetnek, and A. Heinrich. On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems. *arXiv:1303.2875*, 2013.
- [8] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster. Minimization of Tikhonov functionals in Banach spaces. *Abstract and Applied Analysis*, 2008.
- [9] D. W. Boyd. The power method for l^p norms. *Linear Algebra and its Applications*, 9:95–101, 1974.

- [10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
- [11] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 40(1):120–145, May 2011.
- [12] I. Cioranescu. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*. Mathematics and Its Applications. Springer, 1990.
- [13] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.
- [14] O. Hanner. On the uniform convexity of L^p and l^p . *Ark. Mat.*, 3:239–244, 1956.
- [15] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective. *SIAM J. Img. Sci.*, 5(1):119–149, Jan. 2012.
- [16] T. Hohage and F. Werner. Iteratively regularized Newton-type methods with general data misfit functionals and applications to Poisson data. *Numerische Mathematik*, 123(4):745–779, 2013.
- [17] B. Kaltenbacher and B. Hofmann. Convergence rates for the iteratively regularized Gauss-Newton method in Banach spaces. *Inverse Problems*, 26:035007, 2010.
- [18] S. Kamimura and W. Takahashi. Strong convergence of a proximal-type algorithm in a Banach space. *SIAM J. on Optimization*, 13(3):938–945, 2002.
- [19] D. A. Lorenz and T. Pock. An accelerated forward-backward algorithm for monotone inclusions. *CoRR*, abs/1403.3522, 2014.
- [20] D. M. Paganin. *Coherent X-ray Optics*. Oxford Series on Synchrotron Radiation. New York: Oxford University Press, 2006.
- [21] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV '11*, pages 1762–1769, 2011.
- [22] R. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33(1):209–216, 1970.
- [23] R. Rockafellar. *Convex Analysis*. Princeton University Press, 1997.
- [24] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, Nov. 1992.
- [25] F. Schöpfer, A. K. Louis, and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22(1):311329, 2006.
- [26] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. Kazimierski. *Regularization Methods in Banach Spaces*, volume 10 of *Radon Series on Computational and Applied Mathematics*. De Gruyter, 2012.

- [27] R. Stück, M. Burger, and T. Hohage. The iteratively regularized Gauss-Newton method with convex constraints and applications in 4Pi-microscopy. *Inverse Problems*, 28:015012, 2012.
- [28] M. Taylor. *Partial Differential Equations: Nonlinear Equations*, volume 3. Springer, New York, 1996.
- [29] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.
- [30] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681, 2013.
- [31] T. Valkonen. A primal-dual hybrid gradient method for nonlinear operators with applications to MRI. *Inverse Problems*, 30(5):055012, 2014.
- [32] S. Villa, S. Salzo, L. Baldassarre, and A. Verri. Accelerated and inexact forward-backward algorithms. *SIAM Journal on Optimization*, 23(3):16071633, 2013.
- [33] F. Werner and T. Hohage. Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data. *Inverse Problems*, 28:104004, 2012.
- [34] Z.-B. Xu. Characteristic inequalities of l^p spaces and their applications. *Acta Math. Sinica*, 32(2):209–218, 1989.
- [35] Z.-B. Xu and G. Roach. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *Journal of Mathematical Analysis and Applications*, 157(1):189–210, 1991.
- [36] C. Zălinescu. *Convex analysis in general vector spaces*. River Edge, NJ : World Scientific, 2002. Includes bibliographical references and index.